

Preparation and evaluation of sufficiently homogeneous and stable reference materials for priority hazardous substances in whole water

Saioa Elordui-Zapatarietxe¹ · Ina Fettig² · Janine Richter² · Rosemarie Philipp² · Guido Vanermen³ · Els Monteyne⁴ · Gerard Boom⁵ · Håkan Emteborg¹

Received: 21 September 2015 / Accepted: 18 December 2015 / Published online: 4 February 2016
© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract We have prepared and evaluated three whole water test materials containing eight polycyclic aromatic hydrocarbons (PAHs), six polybrominated diphenyl ethers (PBDEs) and tributyltin (TBT) with respect to homogeneity and short-term stability. The test materials were used as samples in two inter-laboratory comparisons. The materials were composed of natural mineral water and model suspended particulate matter (SPM) containing the target compounds at ng L⁻¹ levels. The expanded uncertainty of the estimated mass concentrations in the final test materials was obtained by combining contributions from the homogeneity, the stability and the model SPM characterization. The whole water materials were sufficiently homogeneous and stable at 4 °C for their intended use. In total, 12 out of 15 investigated target parameters could be assessed to be present with a relative combined expanded uncertainty below 25 %. The outcome of the two inter-laboratory comparisons confirmed the good quality of the test materials and the level of uncertainties associated with the estimated mass concentrations. These findings are an

important contribution towards the development of whole water matrix reference materials certified for PAH, PBDE and TBT in support of the Water Framework Directive of the European Union.

Keywords Homogeneity · Stability · Uncertainty · Whole water · Reference materials · PAHs · PBDEs · TBT · Water Framework Directive

Introduction

The Water Framework Directive, WFD, 2000/60/EC [1] and amending Directives establish the legal framework for the protection of European water bodies. According to Directive 2008/105/EC and the more recent amending Directive 2013/39/EC, priority hazardous substances must be monitored in water and biota by the Member States to ensure that the environmental quality standards (EQS) are met [2, 3]. Polycyclic aromatic hydrocarbons (PAHs), polybrominated diphenyl ethers (PBDEs) and tributyltin (TBT) are among the priority hazardous substances due to their toxicity and widespread environmental occurrence. Laboratories face several challenges when monitoring these pollutants in whole water. The EQS for some pollutants are below ng L⁻¹ level, and the samples have to be analysed without the removal of the suspended particulate matter (SPM). Also, the target analytes are strongly bound to the particulate phase in the water, thus complicating easy liberation and extraction [4, 5].

Directive 2009/90/EC obliges monitoring laboratories to use reference materials (RMs) and certified reference materials (CRMs) if available. The use of (C)RMs ensures the accuracy of measurement results and provides comparability and traceability of the measurement results [6].

✉ Håkan Emteborg
Hakan.Emteborg@ec.europa.eu

¹ Institute for Reference Materials and Measurements (IRMM), European Commission, Joint Research Centre (JRC), Retieseweg 111, 2440 Geel, Belgium

² Bundesanstalt für Materialforschung und -prüfung (BAM), Richard-Willstätter-Str. 11, 12489 Berlin, Germany

³ Flemish Institute for Technological Research (VITO NV), Boeretang 200, 2400 Mol, Belgium

⁴ Koninklijk Belgisch Instituut voor Natuurwetenschappen, Linierregimentsplein 3 and 23, 8400 Ostend, Belgium

⁵ TNO, Princetonlaan 6, 3584 CB Utrecht, The Netherlands

RMs also play a pivotal role in proficiency testing (PT) schemes although it might not be directly apparent. The PT samples have to be RMs by definition since they must be sufficiently homogeneous and stable during the time needed for the PT round [7, 8]. The definition for a reference material as given in ISO Guide 30 reads: a material that is sufficiently homogenous and stable with respect to one or more specified properties, which has been established to be fit for its intended use in a measurement process [8]. RMs for PAHs, PBDEs and TBT in whole water that could support the implementation of the WFD are currently not available. Moreover, hydrophobic organic compounds should be present at very low levels in an aqueous matrix where they are associated with natural colloids and SPM in order to mimic natural conditions [9, 10]. The preparation of RMs and of CRMs, in particular for these pollutants in whole water at relevant concentration levels, is therefore difficult. The main challenge lies in ensuring equivalence between the test items (homogeneity) and guaranteeing sufficient stability of the target analytes during the intended lifetime of the material.

In a dedicated project (ENV08) of the European Metrology Research Programme (EMRP), several whole water test materials for PAHs, PBDEs and TBT at low ng L^{-1} were developed as described by Elordui-Zapatarietxe et al. [11]. At the end of the project, these materials were used in two inter-laboratory comparisons (ILCs). In the first ILC (ILC-1), National Metrology Institutes and national laboratories participated to test their analytical capabilities [12]. In the second ILC (ILC-2), similar whole water samples were prepared to validate new and amended standard methods developed under CEN mandate M/424/TC230 in support to the WFD [13–16]. Consequently, the developed materials had to be homogeneous and stable during the lifetime of these ILCs. In this paper, we present full uncertainty budgets for the estimated mass concentrations of the target parameters in whole water samples. The robust mean and means resulting from the two ILCs were also compared with the estimated mass concentrations.

Materials and methods

Developed test materials

Three different test materials were prepared, i.e. one per analyte group (PAHs, PBDEs and TBT), using natural mineral water and model SPM incorporating tenaciously bound priority substances. The preparation method of the test materials, origin and characterization of the model SPM has been described in detail elsewhere [11]. In another paper, Elordui-Zapatarietxe et al. [17] also

investigated the analyte–bottle interactions before the preparation of such test materials. Following this work, all the samples were subsequently prepared and contained in 1-L amber glass bottles with screw caps coated with polytetrafluoroethylene (PTFE) inserts. In total, 15 priority substances listed in the WFD [2, 3] were assessed, namely eight PAHs [naphthalene (N), anthracene (A), fluoranthene (F), benzo(b)fluoranthene (B(b)F), benzo(k)fluoranthene (B(k)F), benzo(a)pyrene (B(a)P), indeno(1,2,3-cd)pyrene (I) and benzo(ghi)perylene (B(ghi)P)], six PBDEs (BDE28, BDE47, BDE99, BDE100, BDE153 and BDE154) and TBT.

The combined standard uncertainty, u , of the mass concentration of each analyte in the different test materials was estimated using Eq. 1:

$$u = \sqrt{u_{\text{bb}}^2 + u_{\text{sts}}^2 + u_{\text{char}}^2} \quad (1)$$

where u_{bb} is the contribution from between-bottle heterogeneity, u_{sts} is the uncertainty arising from the short-term stability (STS), and u_{char} is the uncertainty related to the characterization of the model SPM [11]. A combined relative expanded uncertainty ($k = 2$) of <25 % of the estimated mass concentration was considered achievable but challenging, considering the low concentrations of the priority substances. For the sake of comparison, Member States are required to have a performance of uncertainty of measurements 50 % or below ($k = 2$) [6].

Homogeneity studies

Equivalence between all units produced in a candidate sample batch is a requirement for any RM. Consequently, between-bottle variability of the target compounds has to be quantified and assessed as outlined in ISO Guide 35 [18].

For each type of test material, samples with SPM loads from 7.5 mg L^{-1} to 200 mg L^{-1} were prepared due to requirements of the ILCs (full details in [12, 14–16]). The homogeneity of the three types of test materials was tested at the lowest of the SPM loads used in the ILCs: 20 mg L^{-1} SPM for PAHs and 20 mg L^{-1} for PBDE materials and about 7.5 mg L^{-1} SPM for the TBT material. The lowest amount of SPM was assumed to be associated with the highest degree of variability coming from the preparation step, as well as the lowest concentration in the samples also resulting in a higher variability in the measurement step.

Homogeneity was assessed by analysing five independent units per analyte group measured by gas chromatography coupled with mass spectrometry (GC–MS) for PAHs (Agilent 7890A Agilent Technologies, Santa Clara, CA, USA), gas chromatography coupled with high-resolution mass spectrometry (GC–HRMS) for PBDEs (Waters Autospec Premier, Waters Corp., Milford, MA,

USA) and gas chromatography coupled with inductively coupled plasma mass spectrometry (GC–ICPMS) for TBT (Agilent 7500cx, Agilent Technologies). Specific details about the analytical methodology for PAH and PBDE assessing homogeneity and stability as well as data obtained during ILC-2 are described in [14, 15].

Measurements were performed under repeatability conditions. Each sample unit was vigorously shaken to re-suspend the added model SPM, extracted and analysed without subsampling. Homogeneity assessment based on analysis of variance (ANOVA) to calculate between- and within-bottle heterogeneity could therefore not be applied [18]. Between-bottle heterogeneity (s_{bb}) was consequently quantified by using the relative standard deviation of the mean.

The obtained data were investigated for outlying values using a Grubbs test at 99 % confidence level. Normal distribution of the data was also checked using normal probability plots and histograms (with the limitation of $n = 5$), ensuring that the standard deviation of the obtained data was an appropriate criterion for between-bottle heterogeneity.

Stability studies

Stability studies are essential to check that the levels of the target parameters are maintained for a specific period. Long-term stability (LTS) studies provide information about suitable storage conditions of an RM, and STS information allows establishing the necessary dispatch conditions. As previously reported by Elordui-Zapaterietxe et al. [11], a significant degradation of PAHs and TBT in γ -irradiated samples took place when stored at 60 °C for 4 weeks. Such a high temperature is typically applied to check for analyte stability at the maximum temperature that is anticipated to prevail during shipment. For this reason, it was decided not to perform further studies at 60 °C with new samples (that were not irradiated).

Consequently, the STS of the samples was checked at low (4 °C) and medium temperatures (18 °C) for 4 weeks. The selected temperatures correspond to the preservation of materials in a fridge before analysis (4 °C). The shipment conditions using insulating materials and overnight courier were assumed not to exceed 18 °C. The selected time frame was also the maximum period foreseen from the preparation of the materials until their analysis by the participating laboratories in the ILCs. Longer stability of the material was not required for its intended use in this study, and as a consequence, we did not perform a LTS study. Stability at 18 °C was checked using an isochronous scheme [19]. Samples were kept at 18 °C and then moved to a reference temperature (4 °C) after 0, 1 and 4 weeks. At the end of the evaluation period, all samples were

processed on the same day under repeatability conditions. Two bottles were measured per time point, analysing the entire content of each bottle as described above for the homogeneity testing. The isochronous design results in a higher significance of stability data because the results are not masked by data variability coming from between-day variation at the time of measurement.

In ascertaining the stability of the target parameters at 4 °C, an isochronous scheme could not be followed since the test temperature was identical to the reference temperature. Although it was possible to freeze the samples by special handling of the glass bottles in the freezer, storage of reference samples at negative temperature was not considered since 1 out of 10 bottles broke upon freezing. For this test temperature, all samples were kept at 4 °C after preparation and two units were analysed at each time point, at 0, 1 and 4 weeks. The main disadvantage of this method is that measurements are taken under conditions of intermediate precision. As a result, the increased analytical variation can lead to a higher uncertainty contribution from the STS [20].

All the samples were analysed using GC–MS (PAHs) and GC–HRMS (PBDEs) and GC–ICPMS (TBT) as described above. The data were screened for outliers using a Grubbs test at 99 % confidence level. Linear regression as a function of time was performed to check for statistically significant trends indicating degradation of the analytes. The slopes were tested for significance using a two-tailed t test for a significant level of $\alpha = 0.01$ (99 % confidence interval) [20]. The data sets without significant trends were used to estimate the uncertainty of STS.

Results and discussion

Homogeneity

Between-bottle heterogeneity was evaluated for the three different test materials (Table 1). The data sets were tested for consistency using a Grubbs outlier test at a 99 % confidence level. No outlying values were found. Individual data showed normal or at least unimodal distribution in all cases.

The relative between-bottle heterogeneity (s_{bb}) was below 9 % for all the target parameters with PBDEs having the highest s_{bb} values. The value of s_{bb} is then used as u_{bb} in Eq. 1. It has to be taken into account that this sample type had analyte concentration levels in the pg L^{-1} range, and as a consequence, a larger influence deriving from the analytical method variability should be expected.

The results clearly demonstrate that the target parameters in these types of test material are sufficiently homogeneous to be used as test items in ILCs.

Table 1 Between-bottle heterogeneity given as relative standard uncertainty u_{bb} for PAHs, PBDEs and TBT in whole water samples

Compound	SPM (mg L ⁻¹)	Concentration (ng L ⁻¹)	u_{bb} (%)
PAHs	20		
N		24	4.1
A		11	2.2
F		190	2.6
B(b)F		62	2.0
B(k)F		46	4.8
B(a)P		47	2.2
I		52	1.6
B(ghi)P		61	2.4
PBDEs	20		
BDE28		3.4 ^a	8.7
BDE47		0.26	7.8
BDE99		0.61	3.7
BDE100		0.09	3.5
BDE153		0.12	4.7
BDE154		0.06	4.6
TBT	7.5	3.7 ^b	3.1

Abbreviations are explained in the main text

^a Concentration of BDE28 is given in pg L⁻¹

^b Concentration reported for cation

Stability

Results obtained from each temperature and time point were evaluated separately (Table 2). One outlier was found among the target compounds, namely for BDE28. As no technical reason was found for the exclusion of this data point, it was retained for evaluation. The slope of the regression line of the mass concentration versus storage time was checked for statistical significance ($\alpha = 0.01$) to assess the stability of the target compounds in the three materials. TBT did not show a statistically significant trend neither at 4 °C nor at 18 °C. For PBDEs, all congeners were stable at 4 °C, while half of the congeners showed a statistically significant positive trend at 18 °C. All PAHs, except naphthalene, displayed a positive, statistically significant trend of the slope at 4 °C. A positive trend suggests formation of the target analytes which is completely unrealistic. The positive trend is caused by an analytical artefact since the PAH concentration measured in the sample at time zero was too low. Stability samples kept at 4 °C measured after 1 month still contained estimated concentrations very similar to the added amount. This is a clear evidence of stability for PAHs in these kinds of samples. On the other hand, no significant trend was observed for most of the PAHs at 18 °C. The higher variability of the analytical data results in a higher uncertainty of the measurement results. It was decided to

preserve the samples at 4 °C immediately after preparation. The dispatch was performed the following day using overnight couriers with immediate storage at 4 °C upon arrival in the laboratories. In total, more than 50 shipments were made, and in all cases except two, the samples were delivered the following day. In this way, the transport had a negligible impact on the levels of the target parameters. Nevertheless, an uncertainty contribution for a transport time of 1 week was finally included in the expanded uncertainty. To this end, data sets corresponding to 4 °C were used for PBDEs and TBT since this was the sample storage temperature applied for both ILCs. For the PAHs, data at 4 °C could not be used due to the positive significant trend mentioned above. As an alternative approach, the 18 °C data set was used to estimate the uncertainty contribution to stability.

ILC-1 was conducted within a period of about 6 weeks. Further proof of stability for all analytes was gathered by Richter et al. [12] during ILC-1. The participating laboratories analysed the samples in a time window from 1 to 6 weeks after preparation. No negative trend was observed when correlating analyte concentration as a function of extraction date. Consequently, water samples analysed up to 40 days after preparation were still very close to the estimated concentrations, thus proving that the samples were sufficiently stable during the whole interlaboratory comparison.

Combined expanded uncertainty of mass concentration of the target parameters in whole water samples

In combining the uncertainties of estimated target analyte concentrations in the final test materials (Eq. 1; Tables 1, 2, 3), between-bottle heterogeneity, STS and characterization of the model SPM were taken into account (Table 4). Taken together, there were sufficient degrees of freedom for the main uncertainty contributions to allow the use of a coverage factor of $k = 2$ (95 % confidence interval). In total, 12 out of 15 target parameters were confirmed to be present with a relative expanded uncertainty below 25 % (except naphthalene, anthracene and benzo(k)fluoranthene). For these three compounds, the high variability coming from the characterization of the model SPM resulted in too high u_{char} values as shown in Table 3, increasing the combined uncertainty. A more rigorous characterization of the model SPMs would most likely resolve this problem.

The between-bottle heterogeneity and the uncertainty contribution from the characterization of the model SPM are the main contributing factors to the uncertainty budget (Fig. 1). The variability introduced by analytical methods plays an important role in both cases since a substantial

Table 2 Summary and results of different parameters used to evaluate STS for the target analytes in whole water samples

Compound	Concentration (ng L ⁻¹)	SPM (mg L ⁻¹)	STS parameters							
			Test temperature = 4 °C				Test temperature = 18 °C			
			Outlier ($\alpha = 0.01$)	Slope (%/week)	Significant slope ($\alpha = 0.01$)	u_{sts} (%)	Outlier ($\alpha = 0.01$)	Slope (%/week)	Significant slope ($\alpha = 0.01$)	u_{sts} (%)
PAHs		96.2								
N	120		No	-3.2	No	1.1	No	-12.9	No	6.4
A	50		No	11.6	Yes	2.3	No	-7.5	Yes	2.1
F	910		No	3.8	Yes	0.4	No	-7.5	No	9.7
B(b)F	300		No	5.3	Yes	0.8	No	1.9	No	2.0
B(k)F	220		No	5.2	Yes	0.8	No	1.9	No	1.5
B(a)P	220		No	5.0	Yes	0.8	No	2.2	No	1.9
I	250		No	5.7	Yes	0.5	No	3.8	No	2.4
B(ghi)P	290		No	6.4	Yes	0.8	No	3.2	No	2.3
PBDEs		200								
BDE28	34 ^a		Yes	-1.8	No	1.0	No	-2.9	No	1.1
BDE47	2.6		No	-0.5	No	1.2	No	-4.9	Yes	1.3
BDE99	6.1		No	-0.8	No	1.0	No	-3.5	No	1.0
BDE100	0.9		No	-1.2	No	0.7	No	-2.7	No	1.2
BDE153	1.2		No	-2.2	No	1.2	No	-5.1	Yes	0.9
BDE154	0.6		No	-2.6	No	1.6	No	-4.4	Yes	0.6
TBT	3.7 ^b	7.5	No	1.4	No	1.3	No	-1.9	No	0.7

Abbreviations are explained in the main text

^a Concentration of BDE 28 is given in pg L⁻¹^b Concentration reported for cation**Table 3** Uncertainty associated with characterization of target analytes directly in the model SPM given as relative standard uncertainty $u_{\text{char}}\%$ for PAHs ($n = 3$), PBDEs ($n = 4$) and TBT ($n = 4$)

Compound	$u_{\text{char}} (\%)$
PAHs	
N	16.2
A	13.0
F	2.5
B(b)F	2.1
B(k)F	22.1
B(a)P	7.4
I	3.6
B(ghi)P	2.1
PBDEs	
BDE28	4.0
BDE47	2.4
BDE99	2.4
BDE100	3.1
BDE153	5.4
BDE154	3.1
TBT	3.1

Abbreviations are explained in the main text

part of the uncertainty comes from measurements rather than the actual heterogeneity or (in)stability of the compounds in the test materials. Lower uncertainties could be obtained by further improvement of validated analytical methods with high repeatability and capability of accurately determining PAHs, PBDEs and TBT in whole water samples at ng L⁻¹ concentrations [20].

Estimated mass concentrations in the final samples

The estimated mass concentrations of target analytes in the final water samples were calculated from the amount of target analyte bound to the model SPM and the mass of model SPM added to each water sample [11, 12, 14–16]. The mass of SPM added to the test samples was found by weighing separate portions of oven-dried model SPM obtained after the test sample preparation. This approach of preparing a reference material is based on the so-called formulation as listed in ISO 13528 [21]. The standard also mentions spiking protocols where the analyte is too readily accessible or too loosely bound in comparison with real samples. In such cases, alternative ways of preparation should be sought to achieve more realistic test samples. For

Table 4 Relative combined expanded uncertainty ($k = 2$) of the estimated mass concentrations of priority substances in whole water samples using Eq. 1

Priority substance	U_{rel} (%)
PAHs	
F	21
B(b)F	7
B(a)P	16
I	9
B(ghi)P	8
PBDEs	
BDE28	19
BDE47	17
BDE99	9
BDE100	9
BDE153	15
BDE154	12
TBT	9

Abbreviations are explained in the main text

the particular samples discussed here, all model SPMs were based on naturally incurred soils and sediments where the incipient target analytes are firmly bound to the matrix as shown by leaching experiments performed by Elordui-Zaparietxe et al. [11].

Following ISO 13528, test materials can be prepared by mixing constituents in specified proportions or by adding a specified proportion of a constituent to a base material [21]. In such a case, the assigned value is obtained by calculation from the masses or volumes used. The limitation of this method (in chemical analysis) is that (1) care is needed to ensure that the base material is effectively free from the added constituent, or that the concentration of the added constituent in the base material is accurately known, (2) the constituents are mixed together homogeneously (where this is required), (3) all sources of error are identified (degradation, adsorption or volatilization, etc.), and (4) there is no chemical transformation between the constituents and the matrix. Considering the way the samples have been prepared for these two interlaboratory comparisons and the limitations listed above, the final target concentrations can be estimated but not calculated as is the case using a pure spike. This is mainly because the physicochemical interactions taking place when adding model SPM to prefilled water bottles are not known in detail.

The contribution of u_{char} was estimated as follows:

$$u_{\text{char}} = \frac{s}{\sqrt{n}} \quad (2)$$

where s is the standard deviation of the results used for the calculation of the average and n is the number of independent data sets [22]. Three or four laboratories performed independent measurements of the model SPMs

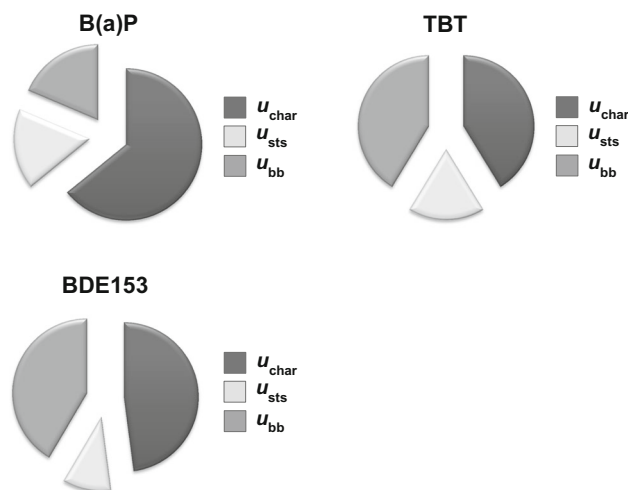


Fig. 1 Uncertainty budgets for selected priority substances in whole water. The relative uncertainty contributions have been normalized to u_{combined} . Abbreviations are explained in the text

and blank SPM [11]. For materials containing both types of SPMs, uncertainties from each SPM were combined calculating the square root of the sum of the squares.

Assessment criteria and results of the two ILCs

As a general assessment of the outcome of the ILCs, the results were checked against a preset criterion of 25 % relative expanded uncertainty. This level of uncertainty was an a priori assumption based on intermediate precision of the measurement methods, knowledge of variation in the sample preparation and initial assessments of stability. It is at the same time a criterion that is twice as stringent as the 50 % relative expanded uncertainty criterion laid down in Directive 2009/90/EC. Admittedly, many of the analytes are present at levels above the EQS in this study, but information available suggests that the 25 % relative expanded uncertainty criterion can be achieved at concentration levels even below the EQS for ILC-1 and data collected during the ENV08 project.

For ILC-1 (ENV08) [12], the output of the interlaboratory comparisons was based on a robust mean as given in ISO 5725-5 using A and S algorithms [23]. For ILC-2 (the CEN M424/TC230 exercise), the reported outcome was based on ISO 5725-2 eliminating outlying values after Cochran and Grubbs tests [14–16, 24].

Figure 2 shows the results of the interlaboratory comparisons compared with the normalized estimated concentrations for the 12 out of 15 investigated priority substances. The dashed horizontal ± 25 % lines show the limits for the applied uncertainty criterion. Each bar displayed by analyte in Fig. 2 consists of 8–13 independent data sets. Each data set in its turn comprised of 2–3

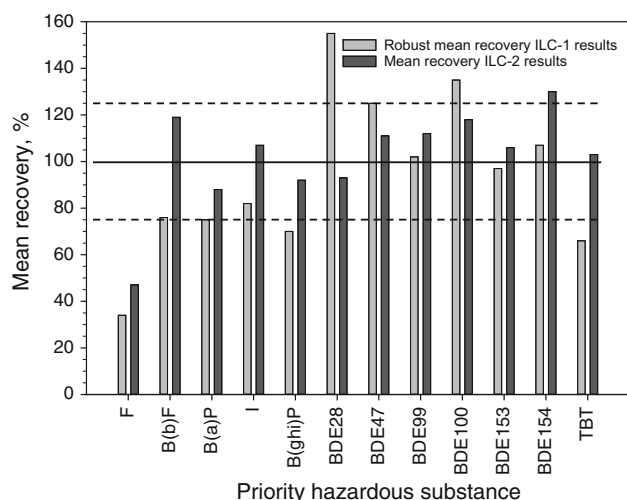


Fig. 2 Comparison of average recovery results from the inter-comparisons to the normalized estimated concentration of each priority substance (100 % line). The dashed lines represent the ± 25 % relative expanded uncertainty criterion. Each bar represents the average recovery of target analyte from 8 to 13 data sets each resulting from analysis of two or three whole water samples. Abbreviations are explained in the text

replicate measurements. Since each measurement is based on the complete extraction of one sample bottle, a total of 156 water samples have been analysed to produce Fig. 2. The robust mean and mean result are directly compared with the 100 % line of the estimated concentrations as a recovery of the added amount. Consequently, the ILC results are within 25 % relative expanded uncertainty of the estimated value, if the mean or robust mean recovery falls within the range from 75 % to 125 %. When applying this 25 % criterion, results for seventeen out of 24 comparisons were within this limit. As can be seen in Fig. 2, for BDE28, the recovery was rather high in ILC-1 (155 %). In this case, the estimated mass concentration was 33 pg L^{-1} and the robust mean was 50 pg L^{-1} from eight data sets [12]. Even though this recovery falls outside the established 25 % criterion, it is still remarkable that eight independent data sets come so close to the estimated concentration despite the extremely low level. Considering the low concentrations and the complexity of the whole water test samples containing substantial amounts of SPM, the final outcome must be regarded as highly satisfactory for both ILCs.

The water samples containing PAHs that were measured in ILC-2 contained an additional blank SPM to obtain the high SPM levels (up to 200 mg L^{-1}) required for validation of the proposed standard method. This blank SPM was obtained by jet milling of a lime-rich agricultural soil prepared in the same way as the other model SPMs [11]. The minor PAH contribution from the blank SPM was also

taken into account when estimating the final PAH concentrations in the test material [14].

With respect to TBT, test samples analysed immediately after preparation by adding slurry to mineral water resulted in recoveries close to 100 %, as discussed by Richter et al. [12]. However, after a few days the recovery values stabilized around 70 % and remained stable as observed during the stability studies and ILC-1. Presently, no explanation can be given for this observation although it shows that the cautions expressed in ISO 13528 are valid for some analytes. For other compounds like the PBDEs, such effects were not observed, and recoveries were high and consistent during the whole exercise.

The best results in ILC-1 (as displayed in the electronic supplementary material in [12]) show that the estimated expanded uncertainties reported here are realistic.

Conclusions

For 12 out of 15 target parameters, the combined expanded uncertainty of the PAHs, PBDEs and TBT concentrations in the test materials was below 25 %. For all studied compounds, the between-bottle heterogeneity and variability coming from the characterization of the model SPM were the main contributors to the combined uncertainty, whereas the uncertainty contribution from stability was smaller. Even though the organic priority substances were present in an aqueous matrix at ultra-trace levels, both the preparation and the analysis of the test samples were successful since in two-thirds of the cases, the interlaboratory comparison means from 8 to 13 data sets were within 25 % of the estimated concentrations. This outcome was possible, thanks to a combination of sufficiently homogeneous and stable test materials and highly capable laboratories applying state-of-the-art analytical techniques that in some cases have been released as official CEN standards thanks to the validation data obtained using these whole water test samples.

Acknowledgments The authors would like to thank the Processing Group at the EC-JRC-IRMM for their constant support during the test material preparation process. The ENV08 project was funded by the European Metrology Research Programme which is jointly funded by the participating countries within EURAMET and the European Union.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Directive 2000/60/EC of the European Parliament and of the Council establishing framework for the Community action in the fields of water policy (2000) Official Journal of European Union L 327:1–73
2. Directive 2008/105/EC on environmental quality standards in the field of water policy (2009) Official Journal of European Union L 348:84–97
3. Directive 2013/39/EC amending Directives 2000/60/EC and 2008/105/EC as regards priority substances in the field of water policy (2013) Official Journal of European Union L 226:1–17
4. Lepom P, Brown B, Hanke G, Loos R, Quevauviller Ph, Wollgast J (2009) Needs for reliable analytical methods for monitoring chemical pollutants in surface water under the European Water Framework Directive. *J Chromatogr A* 1216:302–315
5. Ademollo N, Patrolecco L, Polesello S, Valsecchi S, Wollgast J, Mariani G, Hanke G (2012) The analytical problem of measuring total concentrations of organic pollutants in whole water. *Trends Anal Chem* 36:71–81
6. Directive 2009/90/EC on technical specifications for chemical analysis and monitoring of water status (2008) Official Journal of European Union L 201:36–38
7. Baumeister F, Borchers U, Koch M (2010) PT-WFD: the network of PT providers to support the implementation of the European water framework directive. *Accred Qual Assur* 15:193–198
8. ISO Guide 30 (2015) Reference materials: selected terms and definitions. ISO, Geneva
9. Bercaru O, Ricci M, Ulberth F, Brunori C, Morabito R, Ipolyi I, Sahuquillo A, Rosenberg E (2009) Challenges in preparing water-matrix reference materials for PAHs and pesticides: examples from SWIFT-WFD proficiency-testing schemes. *Trends Anal Chem* 28:1073–1081
10. Ricci M, Kourtchev I, Emons H (2012) Chemical water monitoring under the Water Framework Directive with certified reference materials. *Trends Anal Chem* 36:47–57
11. Elordui-Zapatarietxe S, Fettig I, Philipp R, Gantois F, Lalere B, Swart C, Petrov P, Goenaga-Infante H, Vanermen G, Boom G, Emteborg H (2015) Novel concepts for preparation of reference materials as whole water samples for priority substances at ng L⁻¹ level using model suspended particulate matter and humic acids. *Anal Bioanal Chem* 407:3055–3067
12. Richter J, Elordui-Zapatarietxe S, Emteborg H, Fettig I, Cabillic J, Alasonati E, Gantois F, Swart C, Gokcen T, Tunc M, Binici C, Rodrigues Cea A, Zuliani T, Gonzalez Gago A, Pröfrock D, Nousiainen M, Sawal G, Buzoianu M, Philipp R (2016) An interlaboratory comparison on whole water samples. *Accred Qual Assur*. doi:10.1007/s00769-015-1190-8
13. M/424: mandate for standardisation addressed to CEN for the development or improvement of standards in support of the Water Framework Directive. European Commission, Brussels. <http://ec.europa.eu/growth/tools-databases/mandates>. (Accessed 13 July 2015)
14. European Committee for Standardization, CEN (2015) EN 16691: water quality—determination of selected polycyclic aromatic hydrocarbons (PAH) in whole water samples—method using solid phase extraction (SPE) with SPE-disks combined with gas chromatography mass spectrometry (GC–MS). CEN, Brussels
15. European Committee for Standardization, CEN (2015) EN 16694: water quality—determination of selected polybrominated diphenyl ether (PBDE) in whole water samples—method using solid phase extraction (SPE) with SPE-disks combined with gas chromatography–mass spectrometry (GC–MS). CEN, Brussels
16. European Committee for Standardization, CEN (2015) TS 16692: water quality—determination of tributyltin (TBT) in whole water samples—method using solid phase extraction (SPE) with SPE disks and gas chromatography with triple quadrupole mass spectrometry. CEN, Brussels
17. Elordui-Zapatarietxe S, Fettig I, Richter J, Philipp R, Gantois F, Lalere B, Swart Emteborg H (2015) Interaction of 15 priority substances for water monitoring at ng L⁻¹ levels with glass-, aluminium- and fluorinated-polyethylene bottles for the containment of water reference materials. *Accred Qual Assur* 20:447–455
18. ISO Guide 35 (2006) Reference materials-general and statistical principles for certification. ISO, Geneva
19. Lamberty A, Schimmel H, Pauwels J (1998) The study of the stability of reference materials by isochronous measurements. *Fresenius J Anal Chem* 360:359–361
20. Linsinger TPJ, Pauwels J, van der Veen AMH, Schimmel HG, Lamberty A (2001) Homogeneity and stability of reference materials. *Accred Qual Assur* 6:20–25
21. ISO 13528 (2005) Statistical methods for use in proficiency testing by interlaboratory comparisons. ISO, Geneva
22. van der Veen AMH, Linsinger TPJ, Schimmel H, Lamberty A, Pauwels J (2001) Uncertainty calculations in the certification of reference materials 4. Characterisation and certification. *Accred Qual Assur* 6:290–294
23. ISO 5725–5 (1998) Accuracy (trueness and precision) of measurement methods and results—part 5: alternative methods for the determination of the precision of a standard measurement method. ISO, Geneva
24. ISO 5725–2 (1994) Accuracy (trueness and precision) of measurement methods and results—part 2: basic method for the determination of repeatability and reproducibility of a standard measurement method. ISO, Geneva